



# Using BankIt To Submit 16S rRNAs To GenBank

<https://submit.ncbi.nlm.nih.gov/subs/genbank/>

Submitting microbial 16S RNA sequences using a set of streamlined BankIt web forms

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

## Introduction

Bacterial and archaeal 16S ribosomal RNA (rRNA) sequences are extremely useful reagents for studying the composition of biologically important microbiome communities. The high degree of conservation in key regions of 16S sequences allows amplification with universal PCR primers, while the sequence diversity in other regions allows sample differentiation. A significant fraction of sequences deposited in GenBank are 16S sequences. BLAST searching against this collection can identify the source organisms of the input 16S rRNAs and the taxonomic composition present in the source sample.

## Access

BankIt is a browser form-based tool for submitting relatively small numbers of sequences to GenBank. NCBI has streamlined the BankIt tool to for the submission of 16S rRNAs, reducing the submission to a 9-step process. BankIt is available from the Submission portal (<https://submit.ncbi.nlm.nih.gov>, partially shown above). You need log in to your MyNCBI account (A) and click the “16S rRNA Submission Tool” like (B) to submit your 16S rRNA sequences. Alternatively, you can also use the submission wizard (C) from the Submit page to locate BankIt and other submission tools. Type the data type (for example, 16S) in the text box to see suggested tools (D).

To start a new submission:

- Collect the sequences you want to submit in a plain text file, with sequences in the FASTA format
- Click the “16S rRNA submission Tool” link (B) or the “Start Here” button (D)
- Click the “New submission” button (E) in that page to begin a new submission

## Submit

NCBI collects submissions of data for the world's largest public repository of biological and scientific information.

### Check the Status of Your Submissions ▶

#### QuickStart

You know where you want to go. Select it now!

#### Submission Wizard

Need help figuring out where to start? Try this!

## Submission Wizard

Need help in figuring out where to start? Try this!

The table (F) at the end of the web form lists your existing submissions, their status, and date of last update.

## Steps Needed To Complete A Submission

### Provide your profile

After you click the “New Submission” button, the first screen you need to fill out is the “Submitter” tab (A, shown in part). The server will automatically populate the fields from your existing profile. Make sure you keep the update checkbox checked (B) so the system can keep it up-to-date. Click “Continue” to advance to the next step, the tab in the middle of the page will also update to the next section (C).

**Submission Portal** Home Submissions Objects Groups Templates

Submission: SUB1379283 > GenBank New

**A** Unfinished at the Submitter step  
Delete

Submitter Sequencing technology Sequences Source modifiers References Overview

**Submitter**

\* First name Middle name \* Last name  
John N Doe

\* E-mail (primary) \* E-mail (secondary) **B** Please provide an alternate email address to ensure that messages are received  
john.n.doe@uni.state.edu johndoe100@isp.com

(Other fields were removed for clarity)

Continue ☒ Update my contact information in profile

**C**

Submitter References Sequencing technology

**References**

Sequence authors

\* First name MI \* Last name  
John N Doe

[Add another sequence author](#)

Reference

\* Reference title \* Publication status  
Evaluating the diversity of bacteria in sewage sludge **D** Unpublished In-progress Published

Reference authors

☒ Same as sequence authors ☐ Specify new authors **E**

[Add another reference author](#)

Continue **F**

Submitter References Sequencing technology Sequences Source modifiers Overview

**Sequencing Technology**

Method

\* What methods were used to obtain these sequences? **G**

☐ Sanger dideoxy sequencing  
☒ 454  
☐ Helicos  
☐ Illumina  
☐ IonTorrent  
☐ Pacific Biosciences  
☐ SOLID  
☐ Other

**Assembly State**

Are these sequences:

☐ unassembled sequence reads  
☒ assembled sequences (consisting of two or more sequence reads)

**Assembly Information**

\* Assembly program \* Version or date Delete  
PANDAsq 2.5 **G**

[Add another assembly program](#)

Continue

### Add relevant publications

The Reference tab (C) allows you to add publications to your submission. These references add methods and context to your sequences. When such a publication is not available, the “Unpublished” option (D) allows you to credit colleagues who contributed to the sequencing effort. You can click the “+” link (E) to add input fields for additional authors

### Specify sequencing technology

The “Sequence technology” tab (F) allows you to specify the technology used to obtain the sequences. In addition to the traditional Sanger-dideoxy method, this tab lists several popular next-generation sequencing technologies (G). The 454 option is checked for this example.

You also need to provide the program (G) used to assemble sequence reads into the final submitted form.



## Steps Needed To Complete A Submission (cont.)

### Uploading your sequences

Submitter References Sequencing technology **Sequences** (A) Source modifiers Overview

**Sequences**

Release date

\* When this submission should be released to the public (B)

☐ Release immediately following curation

☒ Release on specified date (not viewable until this date or the release of linked data, whichever is first)

\* Release date (YYYY-MM-DD) (C)

2017-06-01

**Chimera check**

\* Did you check and remove low-quality and chimeric sequences from your FASTA file prior to preparing this submission? (C)

☒ Yes

☐ No

Please provide the name and version of the chimera checking program.  
BLAST alone is not sufficient as a chimera checking program.

Program Name Version

DECIPHER 1.16.1

**Cultured or Uncultured**

\* Bacterial/Archaeal Sequences: How were they obtained? (E)

☒ Pure-cultured strain(s) (axenic culture(s) containing only one microbial species each)

☐ Uncultured, bulk environmental DNA (PCR-amplified directly from environment)

**Sequences**

\* Upload a FASTA-formatted nucleotide sequence file.

16S rRNA\_txt\_filtered.fsa (uploaded)

**Browse...** (D)

Uploading a new file will replace the current file.

Multiple file uploads are not supported at this time, so all of your sequences  
If you have multiple large files, make a separate submission for each file.  
Do not upload more than 5,000 sequences in one file.

How do I create a FASTA file?

**Continue**

The Sequences tab (A) allows you to provide more information about your sequences and to upload them. You will need to specify the public release date (B) for the sequences you are submitting. You should also provide details on how you checked for sequence chimeras (C) to ensure the accuracy of these submitted sequences. Then, upload the sequence file (D). The example sequences in FASTA format (E) contain organism names in square brackets (F).

```
>seq1 [organism=uncultured bacterium]
TCTCCTACCGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCCTGATGCAGCCATGCCCGCT
GTGTGAAGAAGGCTTTCGGGTTGTAAAGCACCTTCAGCGAGGAGGAAAGGGTNTTGTCTTAATACGTAATA
TCTGTGACGTTACTCGCAGAAGAAGCACCGGCTAATCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGC
AAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCTAGGTGGTTAATTAAGTCAGATGTGAAAGCCCGAG
GGCTCAACCTTGGAACTGCATTGAACTGGTTAACTAGAGTTNTGTAGAGGGTGGTGAATTTTCAGGTG
TAGCGGTGAAATGCGTAGAGATCTGAAGGAATACCACTGGCGAAGGCGGCCACCTGGACAGTAAGTGACA
CTGAGGCGCGAAGGCGTGGGGAGCAACGGGATTAGATACCCCGGTAGTCCACGCAGTAACGATGTCTA
TTAGAAGTTTGTGGCTATATGCCGTGGGTCAAAGCTAACGCATAAATAG

>seq2 [organism=uncultured bacterium]
GGGGNAGCAGTGGGGAATATTGCACAATGGGGGAAACCCCTGATGCAGCCATGCCCGCTGTGTGAAGAAG
GCTTTCGGGTTGTAAAGCACCTTCAGCGAGGAGGAAAGGGTGTGCTTAATACGTAACATCTGTGACGTT
ACTCGCAGAAGAAGCACCGGCTAATCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGAACGGCTTAATC
GGAATTACTGGGCGTAAAGCGCGCTAGGTGGTTAATTAAGTCAGATGTGAAAGCCCGAGGCTCAACCTT
GGAAGTGCATTTGAACTGGTTAACTAGAGTTCTGTAGAGGGTGGTGAATTTTCAGGTGTAGCGGTGAAA
TGCGTAGAGATCTGAAGGAATACCACTGGCGAAGGCGGCCACCTGGACAGTAAGTGACACTGAGGCGCGA
AGGCGTGGGGAGCAACGGGATTAGATACCCCGGTAGTCCACGCAGTAACGATGTCTATTAGAAGTTTG
TGGCTATATGCCGTGGGTCAAAGCTAACGCATAA
```

Note that the 16S rRNA BankIt has a limit of a single file with up to 5000 sequences. You will need to break up larger number of sequences into smaller files and submit each separately.

Submitter References Sequencing technology Sequences **Source modifiers** (G) Overview

**Source Modifiers**

1. Required information includes Sequence\_ID, Organism, and strain.

2. Download Source Modifier template with any source information provided so far.

3. You may edit this table in Microsoft Excel or any other editor.

4. The file must be saved as text (tab-delimited).

5. Upload Source Modifiers file.

6. Click Continue to validate the information. If there are warnings or errors, review and correct the information where appropriate and click continue.

More help on source modifiers (J)

The file format is a tab-delimited text file. Shown below is an example Source Modifiers file:

Sequence_ID	Organism	isolation-source	strain	Collection-date	Country
Seq1	Lactobacillus acidophilus	yogurt	ABC1	30-Nov-2012	Iceland
Seq2	Bifidobacteriales bacterium	yogurt	CBS 123	Feb-2003	Greece
Seq3	Bifidobacterium sp.	dairy drink	DEF2	2009	Greece

Upload a Source Modifiers file.

**Browse...** No file selected.

### Add source modifier

Feature annotation provides additional metadata adding great value to submitted sequences. For 16S rRNA sequences the most important metadata are the isolation source and taxonomic identification. The "Source modifier" tab (G) allows you to provide biological source and taxonomic information to sequences you are submitting.

You need to upload a metadata file for your sequences in the form of tab-delimited source modifier file through the "Browse" button (H). You can download a template (I) to model your input. See the help page (J) for more details.

Sequence_ID	Isolation_source	Clone	Country	Host	Collection_date	Lat_Lon
seq1	host hemolymph	1611_16_117	Germany		Crassostrea gigas 21-May-2013	55.0289 N 8.4342 E
seq2	host hemolymph	1611_6_32	Germany		Crassostrea gigas 21-May-2013	55.0289 N 8.4342 E

## Steps Needed To Complete A Submission (cont.)

### Review what you just provided and submit



[Submitter](#)
[References](#)
[Sequencing technology](#)
[Sequences](#)
[Source modifiers](#)
[Overview](#)

#### Overview

Please review your submission and click Submit.

**Submit**

After submission, your data will be examined using BLAST and other checks. You will be contacted if issues are found. You have requested that your sequence data be released on **Jun 01, 2017**. GenBank accession numbers will be assigned to your submissions and sent to you by email within two working days, unless there are issues with your submission that we must ask you to explain first. If you have any questions or corrections regarding your submissions before you receive these, be sure to refer this Submission ID in your email.

**Submitter**

Submitter: **John N Doe**  
[john.n.doe@uni.state.edu](mailto:john.n.doe@uni.state.edu)  
[johnndoe100@isp.com](mailto:johnndoe100@isp.com)

Institution: **State University**  
 Department: **Biological Sciences**  
 Street: **4900 K Street**  
 City: **Farmington**  
 Postal code: **99999**  
 Country: **United States**

**Sequence authors**

- John N Doe

**References**

Reference title: **Evaluating the bacteria diversit in sewage sludge**  
 Publication status: **unpublished**  
 Authors: **same as sequence authors**

**Sequencing Technology**

Methods: **454**  
 Assembly state: **assembled**  
 Assembly Programs: **PANDAsseq (2.5)**  
 Chimera tool used?: **DECIPHER 1.16.1**

**Uploaded files**

- [16SrRNA.txt](#)
- [srcbtl.txt](#)

**Sequence processing reports**

Text report: **SUB934806 16SrRNA txt sequences report txt.txt**  
 Spreadsheet: **SUB934806 16SrRNA txt sequences report tbl.csv**

**Files that will be used for this submission**

Sequences file(s): **SUB934806 16SrRNA txt fasta filtered.fsa**  
 Source Modifier file(s): **srcbtl txt filtered.src**

**Submit**

After you have provided all the necessary information and uplodated both the se-  
 quence and source modifi-  
 er files, BankIt displays  
 them in a single page for  
 you to review for accuracy  
 (A).

The BankIt tool assigns a submission ID (B) to  
 uniquely identify the submission. Refer to this ID  
 in any correspondence about the submission.  
 Click the "Submit" button (C) to complete the sub-  
 mission, and post your sequences for processing  
 by GenBank. You will receive an email acknowl-  
 edging your submission.

GenBank staff may contact you if they have any  
 questions about your submission. They will send  
 you an email with assigned accessions. Before  
 the release date, GenBank staff will send you an  
 email to remind you of the pending release of  
 your sequences. If you want the release post-  
 poned, you will need to reply and set a new date.

### References

A quick overview of sequence submission to  
 NCBI is available at:  
<https://www.ncbi.nlm.nih.gov/guide/howto/submit-sequence-data/>

Additional information on GenBank and submis-  
 sion is available from the GenBank site at:  
<https://www.ncbi.nlm.nih.gov/genbank/>

A set of Youtube video tutorials on BankIt is at:  
<https://www.youtube.com/watch?v=OZxxsRm0pP4> (part 1)  
<https://www.youtube.com/watch?v=DhYUYJSm2mQ> (part 2)

## Technical Support

You can send technical questions about how to use BankIt or problems encountered to:  
[info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

For questions about a successfully completed submission, you should send them to:  
[gb-admin@ncbi.nlm.nih.gov](mailto:gb-admin@ncbi.nlm.nih.gov)  
 Make sure you include the submission IDs assigned to your submission.